Improvements to Hardware-Accelerated 3D Single Particle Imaging Data Reconstruction

Niteya Shah, Christine Sweeney, and Vinay Ramakrishnaiah

I. INTRODUCTION

Fast analysis of scientific data from X-ray free electron laser (XFEL) experimental facilities is key for supporting real-time decisions that efficiently use these facilities to speed up scientific discovery. Our research shows gains obtained using graphics processing units (GPUs) to accelerate 3D reconstruction of Single Particle Imaging (SPI) X-ray diffraction data. We achieve a 4X speedup over the previous GPU implementation, 50% better image reconstruction resolution, and 485X speedup when calculating resolution compared to the existing implementation. We showcase techniques to optimize per-node computational efficiency, increase scalability and improve the accuracy of SPI by using better algorithms, improving data movement and accesses, reusing data structures and reducing memory fragmentation.

II. BACKGROUND

SPI is a technique for determining the structure of bio-macromolecules using diffraction patterns gathered from directing an X-ray beam at single particles. SPI via free electron lasers (FELs) is fundamental for understanding the time evolution of enzymatic reactions. SPI starts with a randomly generated electron density estimation and iteratively builds a closer estimation of the protein structure.
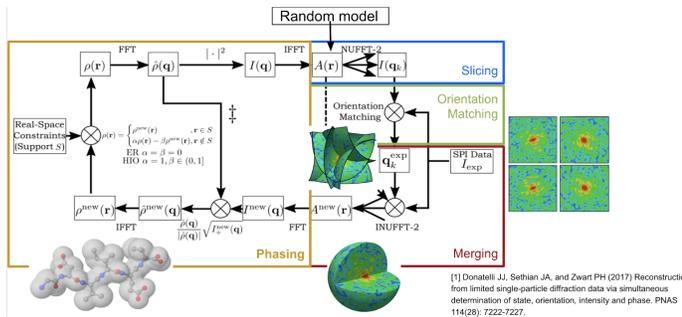

Fig. 1: SPI Workflow

SPI can be split into the following components, shown diagrammatically with Figure 1:

A. Slicing

The slicing step creates M 2D model images from the electron density estimation by applying non-uniform FFTs. SpiniFEL uses the cufinufft [4] library to efficiently calculate the forward (uniform to non-uniform) FFT.

B. Orientation Matching

The orientation matching step finds the closest reference image for each model image generated from slicing. Pairwise Euclidean distance is computed between the reference images and the model images and then the nearest matching reference image is found for each model image.

C. Merging

The merging step combines the reference orientations into a 3D diffraction volume. This is done by solving the linear system using the conjugate gradient (CG) algorithm.

D. Phasing

The phasing step converts the 3D diffraction volume into a molecular structure by applying real-space and reciprocal space constraints on the electron density estimate of the sample.

E. Resolution Calculation

When testing the algorithm, the achieved convergence is calculated by the comparing the normalised cross-correlation coefficient between the generated model and the base model using Fourier shell correlation (FSC).

III. METHODOLOGY

A. Improvements to Slicing and Orientation Matching

The slicing and orientation matching steps are combined to reduce data movement. The Einstein summation convention (einsum) operation for the rotation matrix is precomputed and stored in pinned memory. We then batch the slicing operation to reduce memory utilization, and for every batch, we slice it and then compute a partial orientation match using a GPU-optimized algorithm as follows:

The Euclidean distance of two vectors can be written as

$$(\vec{x} - \vec{y})^2 = ||\vec{x}||^2 + ||\vec{y}||^2 - 2 < \vec{x}, \vec{y} > \tag{1}$$

When implemented across multiple images, this transforms into

$$||(X_{A,n} - Y_{B,n})||^2 = \sum_{i=1}^{n} X_{A,i}^2 \oplus \sum_{i=1}^{n} Y_{B,i}^2 + X_{A,n} \cdot Y_{B,n}^T \tag{2}$$

The outer product is found by broadcasting the squared norm of the images. The matrix product can be computed using the GEMM BLAS routine $C := C - 2 * X * Y^T$. By using the CUBLAS GEMM routine to calculate the matrix product,